'Doing' Responsible AI: Practical Experiences

Duuk Baten
Advisor Responsible Tech (SURF)
13 June 2024

Objective Portrait (2022) – Vera van der Burg & Gijs de Boer

EN L'AN 2000

Villemard, 1910 – At School

Ruining the surprise

1. AI touches everything, it is a *systems-technology*
2. Technologies *shape* our world and we have a *responsibility* for technology
3. *Responsibility* is a practice

SURF

Who am I?

Duuk Baten
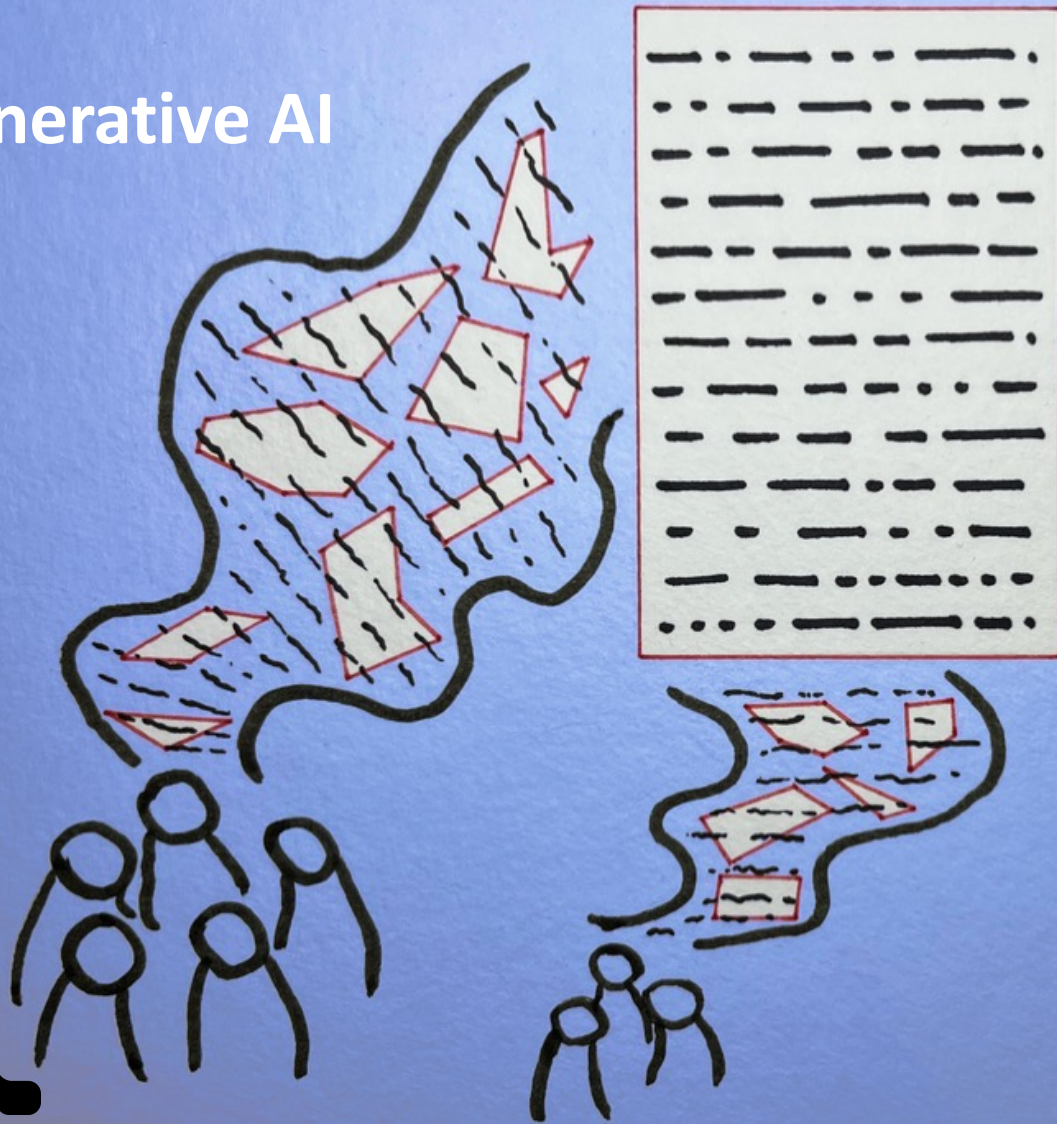**Advisor Responsible Tech**

duuk.baten@surf.nl
SURF Innovation Lab – Program Public Values

- *'Responsible Tech' – how do we use new technology responsibly*

- Member of EU expert group for ethical guidelines for AI in education

- Dutch AI coalition working group Education

- GPT-NL (training a Dutch LLM)

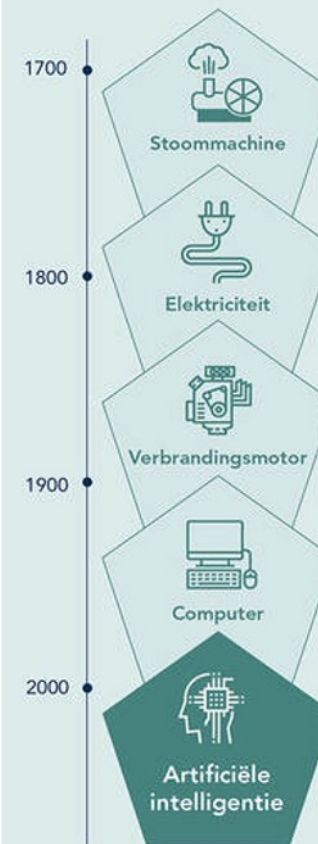- Expedition AI – collaborative approach towards AI in education by SURF & Npuls →

SURF

The state of AI and the modern educational institution.
AI explained in the context of the educational sector

edu.nl/epypq

Generative AI

AI als systeemtechnologie

WRR rapport 105, 2021

TECHNOLOGY

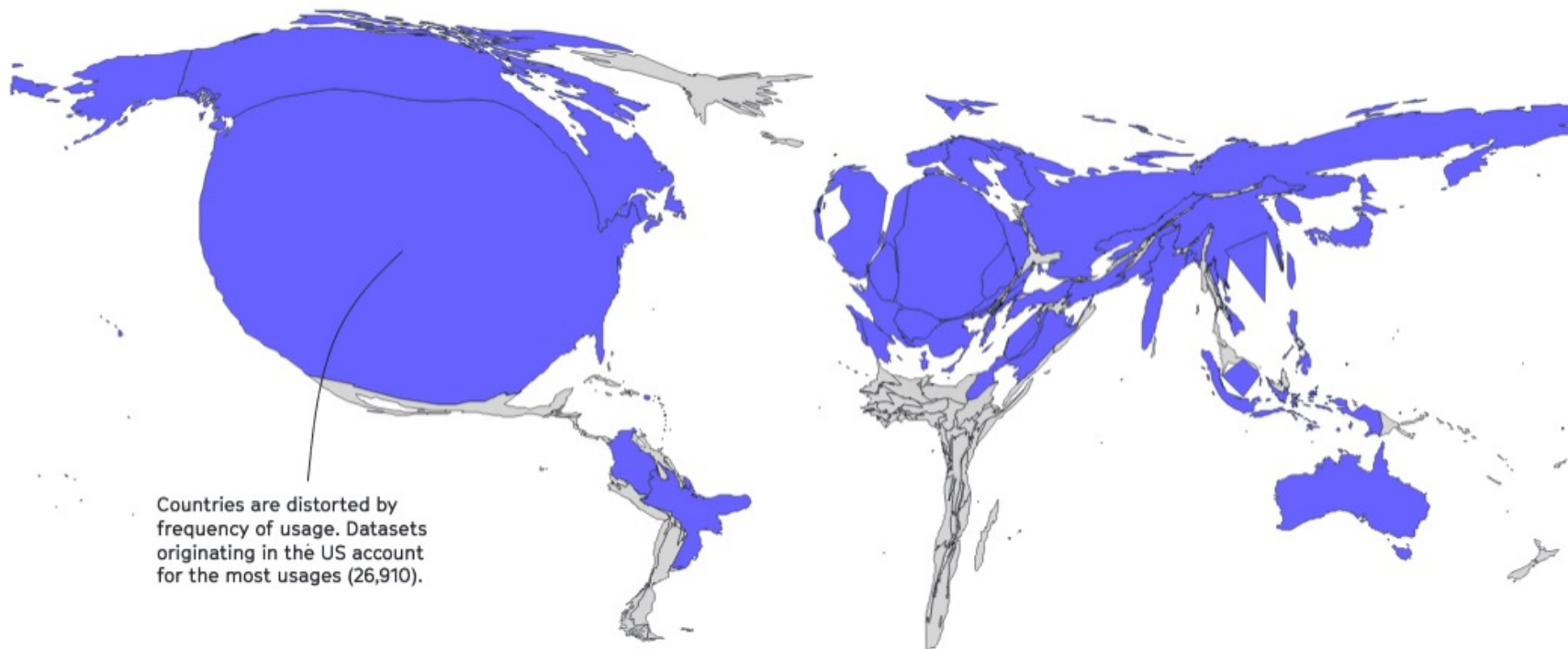DOMAIN

AI touches everything, it is a *systems-technology*

SURF

Technology shapes our view of the world

Ann Street Studio
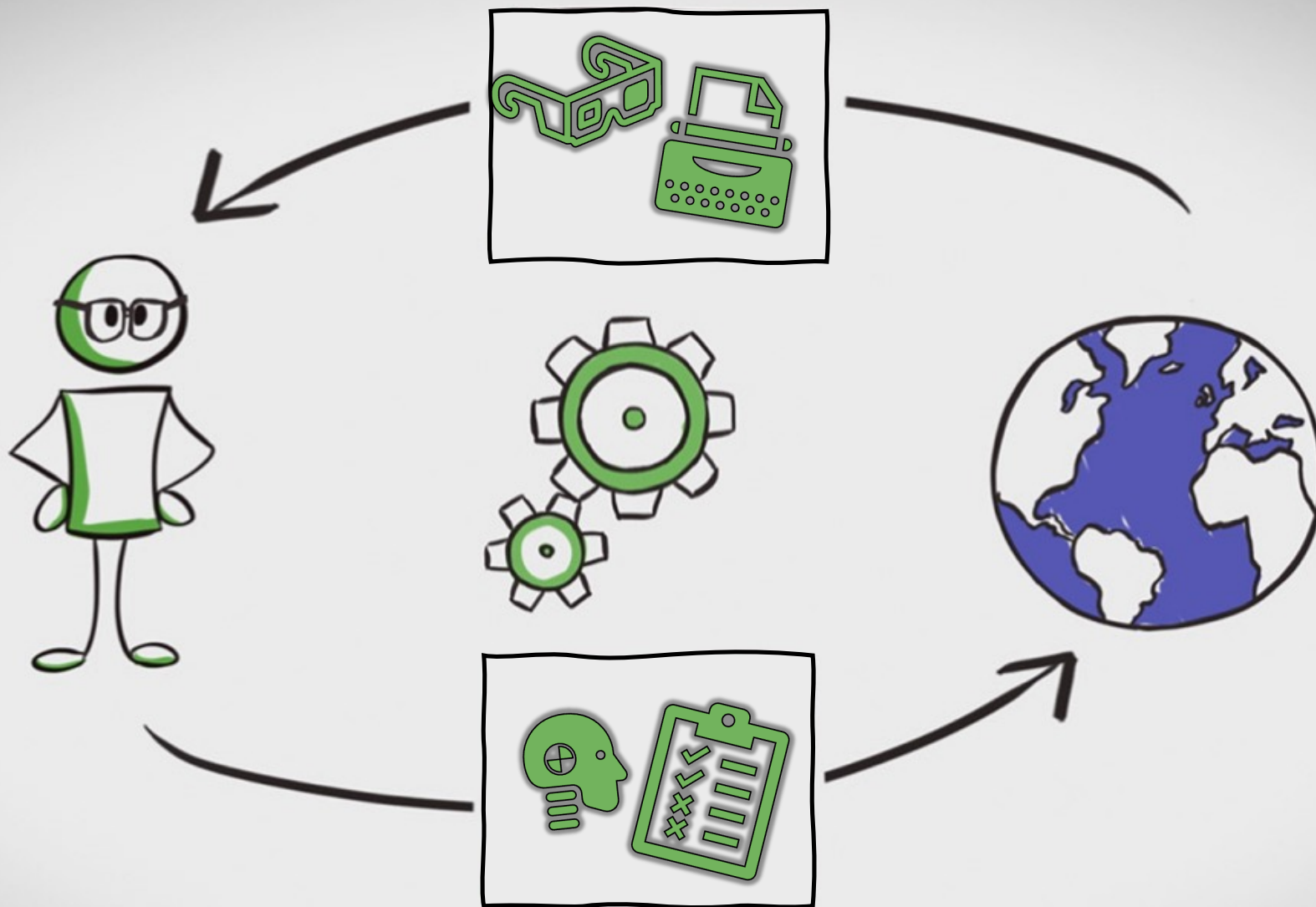
This impact is not neutral

SURF

Frequency of dataset usage by country

● Usage of datasets from here   ● No usage of datasets from here

Countries are distorted by
frequency of usage. Datasets
originating in the US account
for the most usages (26,910).

SURF

ⓘ This map shows how often 1,933 datasets were used (43,140 times) for performance benchmarking across 26,535
different research papers from 2015 to 2020.

A rise in challenges

NEWS | 24 October 2019 | Update 26 October 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Heidi Ledford

Download PDF

Black people with complex medical needs were less likely than equally ill white people to be referred to programmes that provide more personalized care. Credit: Ed Kashi/VII/Redux/eyevine

United Nations | UN News
Global perspective Human stories

Urgent action needed over artificial intelligence risks to human rights

Unsplash/Michael Dziedzic | Artificial intelligence could help to boost the provision of healthcare around the world.

15 September 2021 | Human Rights

iapp

Privacy Perspectives

Digital welfare fraud detection and the Dutch SyRI judgment

Sep 23, 2021   Save This

Frederik Zuiderveen Borgesius
Nonmember Contributor

Marvin van Bekkum
Nonmember Contributor

States worldwide are turning to technology to make the welfare state more efficient and mitigate welfare fraud. In the Netherlands, the state used a digital welfare fraud detection system called Systeem Risico Indicatie. The SyRI was a system that used personal data from different sources and uncovered fraud. In 2020, a Dutch court decided the SyRI legislation was unlawful because it did not comply with the right to privacy under the European Convention of Human Rights. This is among the first time a court invalidated a welfare fraud detection system for breaching the right to privacy. We analyze the judgment and its implications in a full paper; below are some of our main points.

16

Technologies *shape* our world and we have a *responsibility* for technology

SURF

How do you make sure you responsibly engage with new technology?

SURF

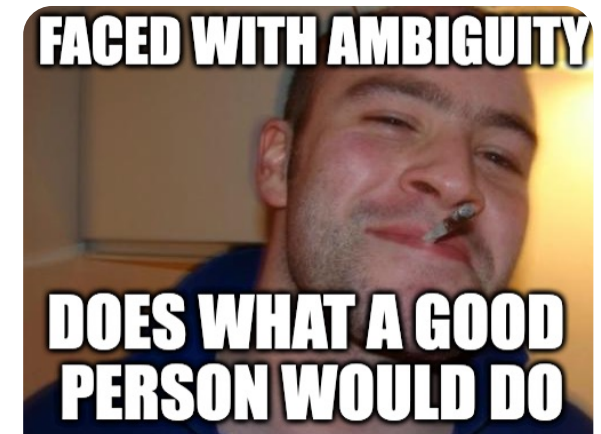# 3 main types of *Western* ethical theory

Duty-based Ethics
What duties do we have towards ourselves and others?

Consequentialism
The morally best choice is that which gives the best end-state.

Virtue Ethics
What does it mean to be a good person, to make the right choices.

SURF

Dealing with Reality

"Culture of accountability"

Helen Nissenbaum (1977) identified four barriers to accountability in digital systems and society:

(1) The problem of many hands

(2) The acceptance of computer bugs as an inherent element of large software systems

(3) Using the computer as scapegoat

(4) Ownership without liability

Responsibility as Practice

SURF
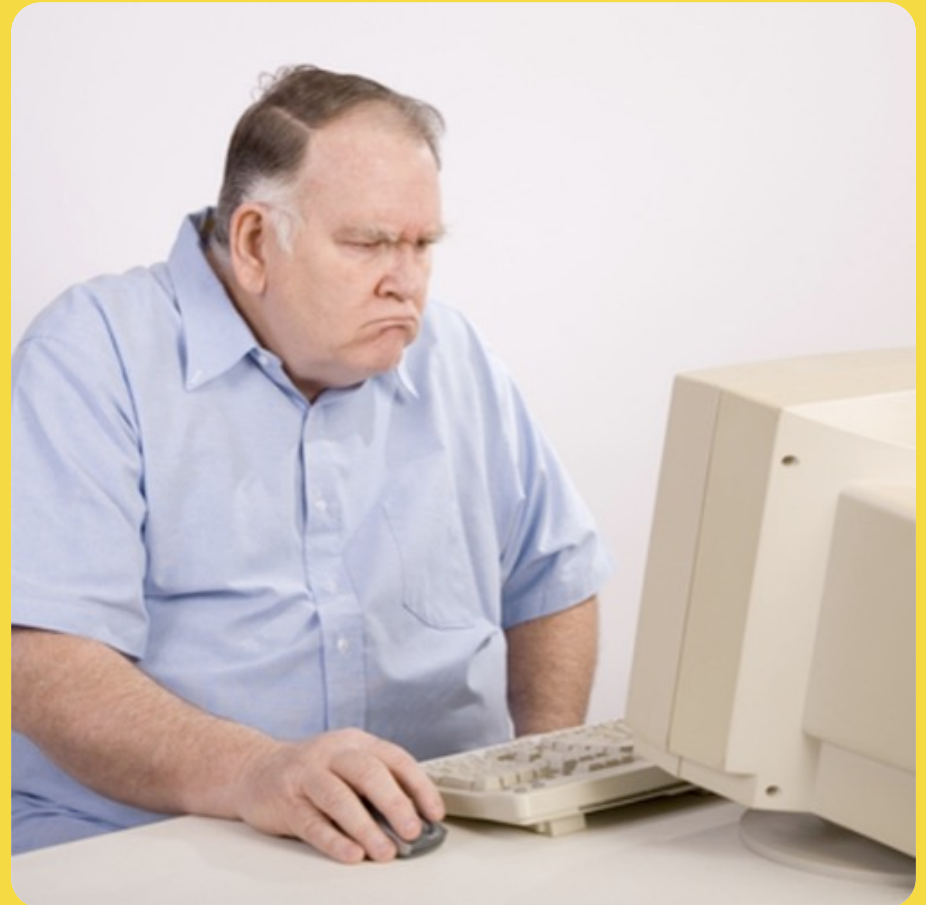
Responsibility as Practice

→ The idea that responsibility is *not just a theoretical concept*, but *a practical skill* that needs to be developed and exercised.

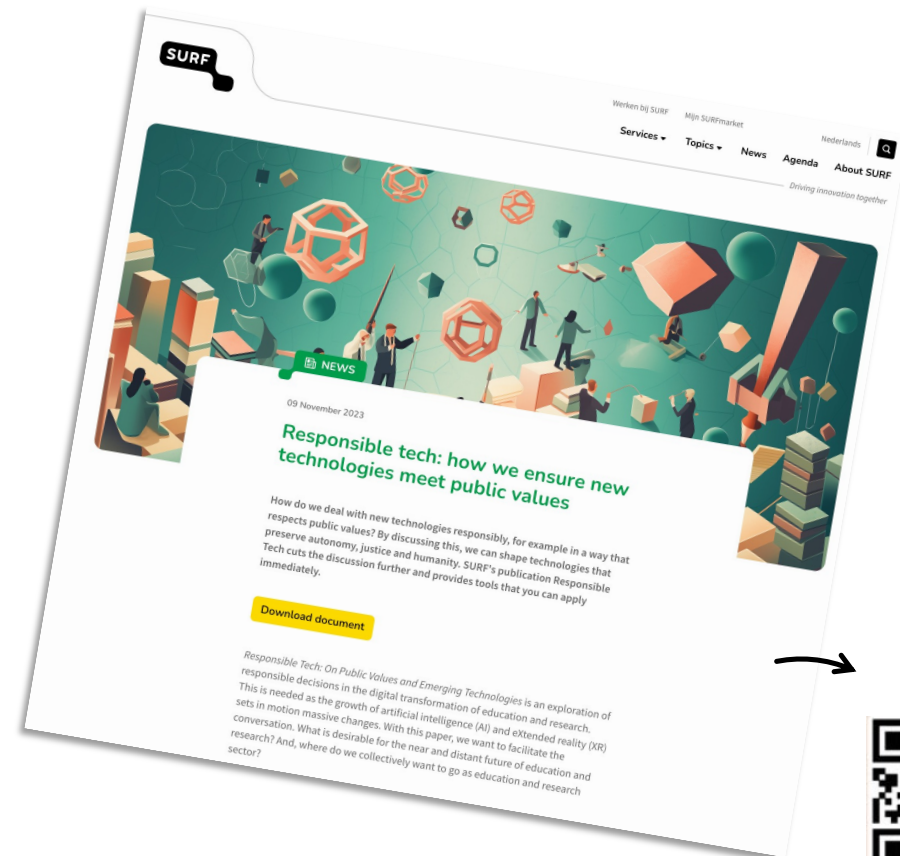SURF

"Ethics is a process, not a destination"

National Academies of Sciences, Engineering, and Medicine (USA)

Responsibility is a practice
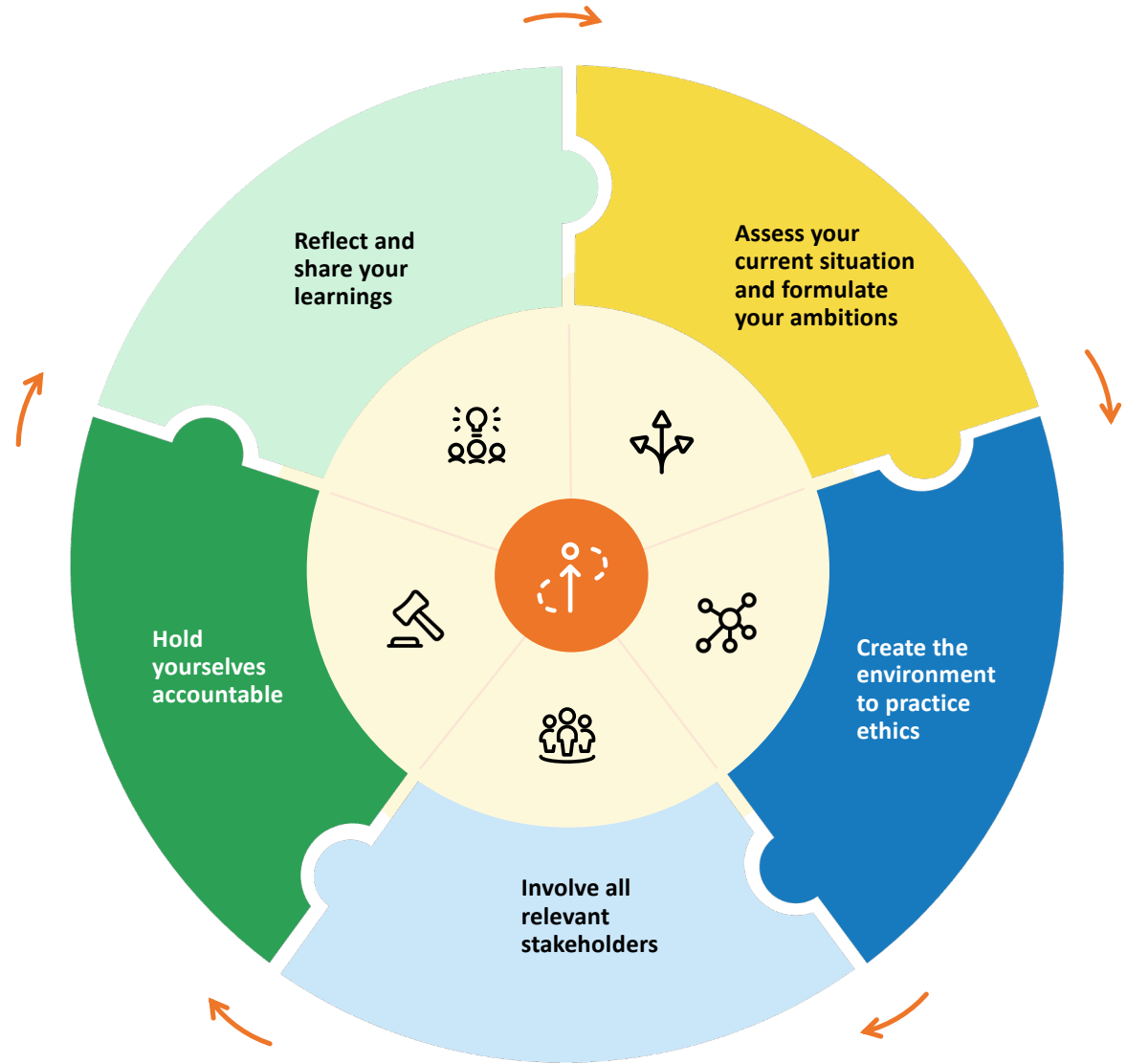
A good start
in 5 steps

We formulated **5** recommendations in our discussion paper

'Responsible tech: how we ensure new technologies meet public values'
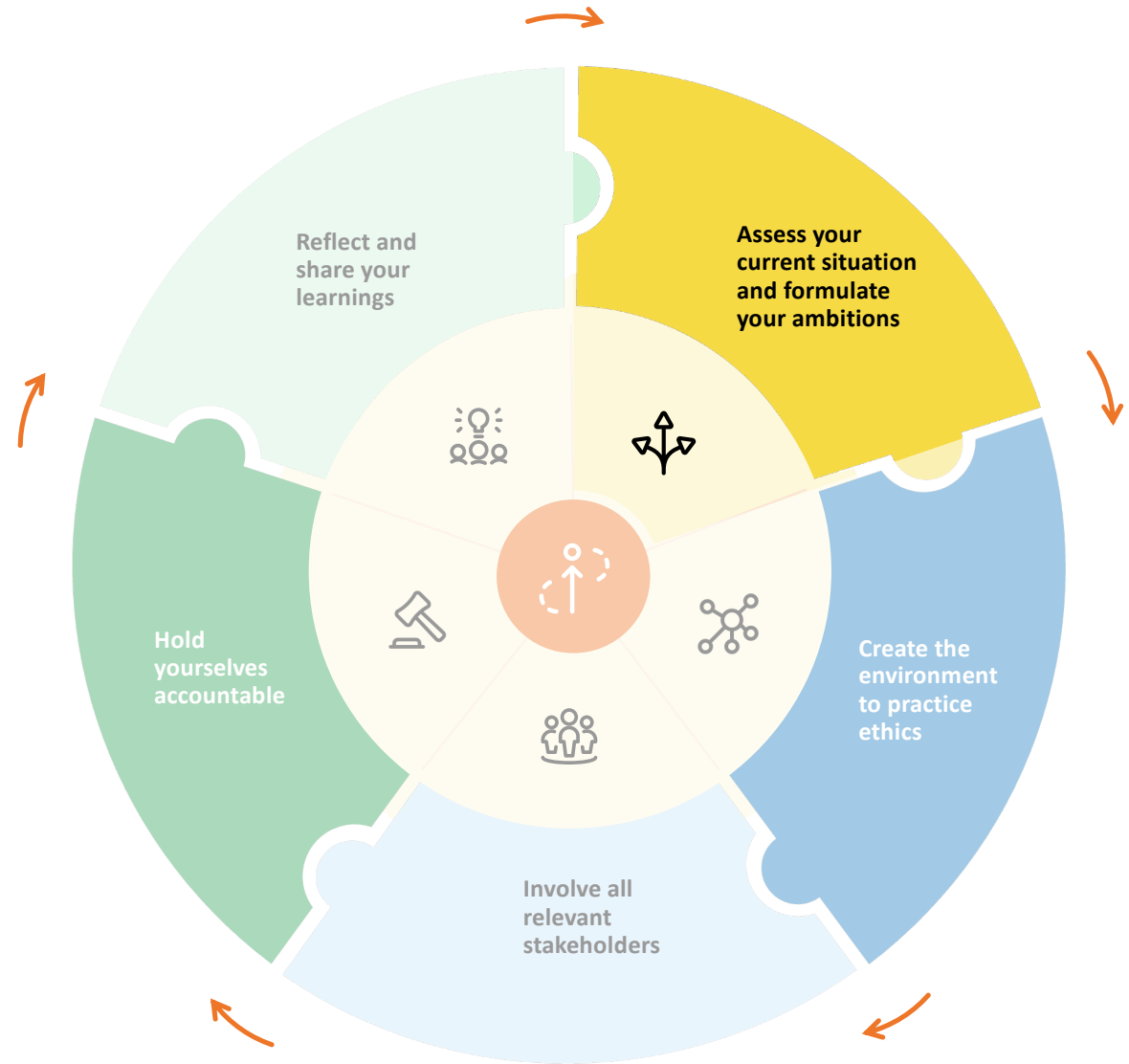
How do we get started?

- Reflect and share your learnings
- Assess your current situation and formulate your ambitions
- Create the environment to practice ethics
- Involve all relevant stakeholders
- Hold yourselves accountable

How do we get started?

- Recognise where you are at. E.g. impact assessments (IAMA) of AI ethics maturity-models.

- Formulate your ambitions. What are you trying to achieve, which values are central?

- Make choices, do you want to use certain technologies?

SURF

Reflect and share your learnings

Assess your current situation and formulate your ambitions

Create the environment to practice ethics

Involve all relevant stakeholders

Hold yourselves accountable

Pilot with AI Ethics Maturity framework (Krijger et al. 2022)

Awareness & Culture
Ethical AI practices are integrated in the organization.

Policy
Is available and implemented throughout organisation.

Governance
Internal procedural ethical checks and balances are in place.

Communication & Training
Internal communication about ethics and training.

Development processes
Working in structured approaches aligned with ethics.

Tooling
Can give insights into monitoring, discussing or improving ethical AI.

SURF

Value Compass for
digital transformation of education

Autonomy

- Freedom of choice
- Privacy
  - Safeguarding of private life and personal data
- Self-determination of pupils and students
- Diversity
  - Freedom of education
- Independence of educational institutions
- Professional autonomy of teachers/lecturers
- Independance of education

Justice

- Accessibility
- Inclusivity
- Equality
  - Equal opportunities
  - Equal treatment
- Integrity
  - Trustworthiness of information, content, data, and systems
  - Transparency
  - Effectiveness
  - Democratic control
  - Sustainability

Humanity

- Social cohesion
- Health, well-being
- Meaningful contact
- Safety
- Personal development
- Respect

Kennisnet SURF

edu.nl/68vd4

How do we get started?

- Recognise where you are at. E.g. impact assessments (IAMA) of AI ethics maturity-models.

- Formulate your ambitions. What are you trying to achieve, which values are central?

- Make choices, do you want to use certain technologies?

SURF

How do we get started?

- Use an iterative approach, try something, make choices, and reflect

- Make room for tensions and discomfort

- Arrange for diverse project-teams, steering-groups, and advisory committees
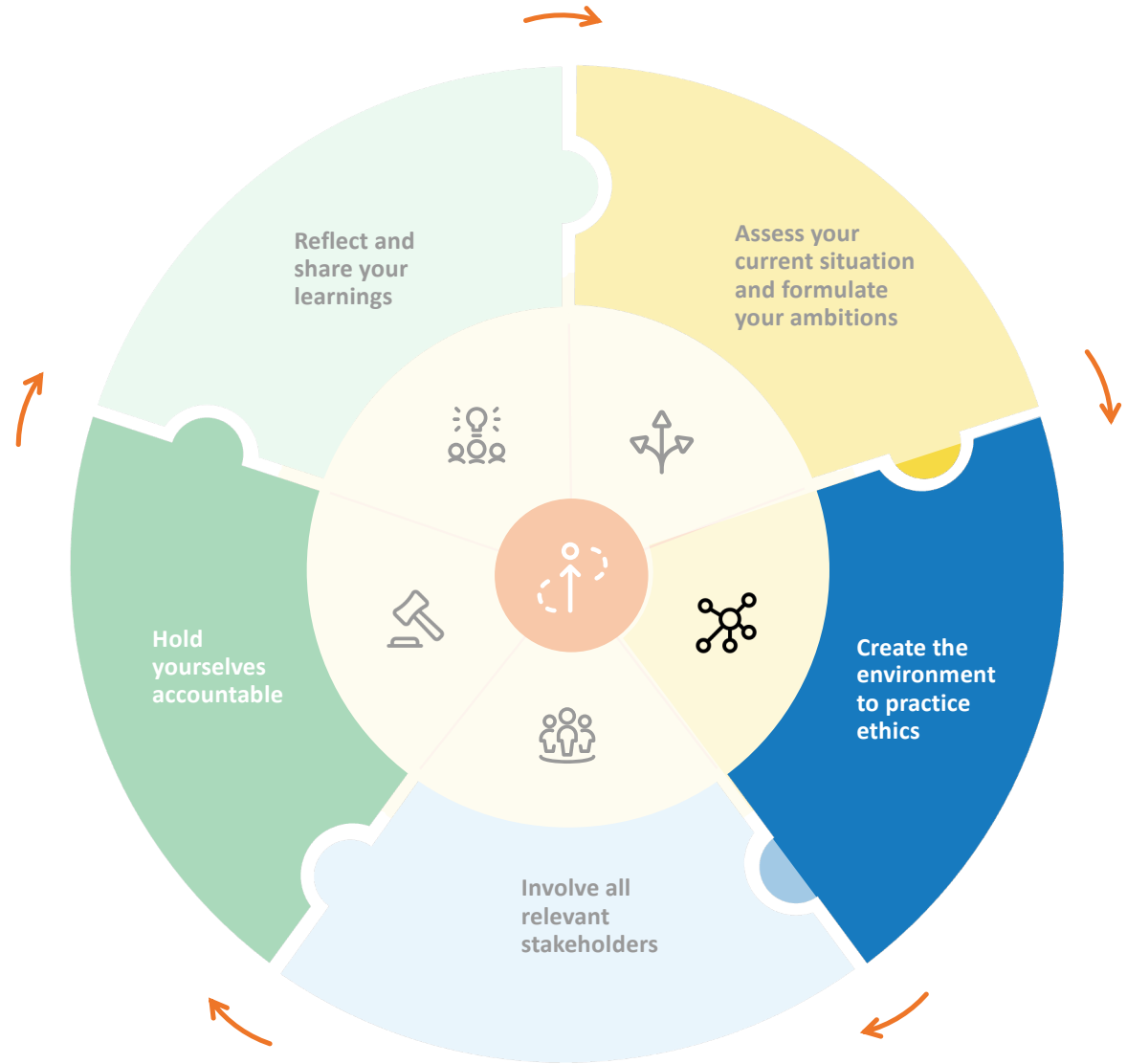
How do we get started?

- Create the needed structure to support taking responsibility (e.g. ethics boards)

- Prioritize transparency around technology (e.g. commitments)

- Take responsibility for the applications you promote
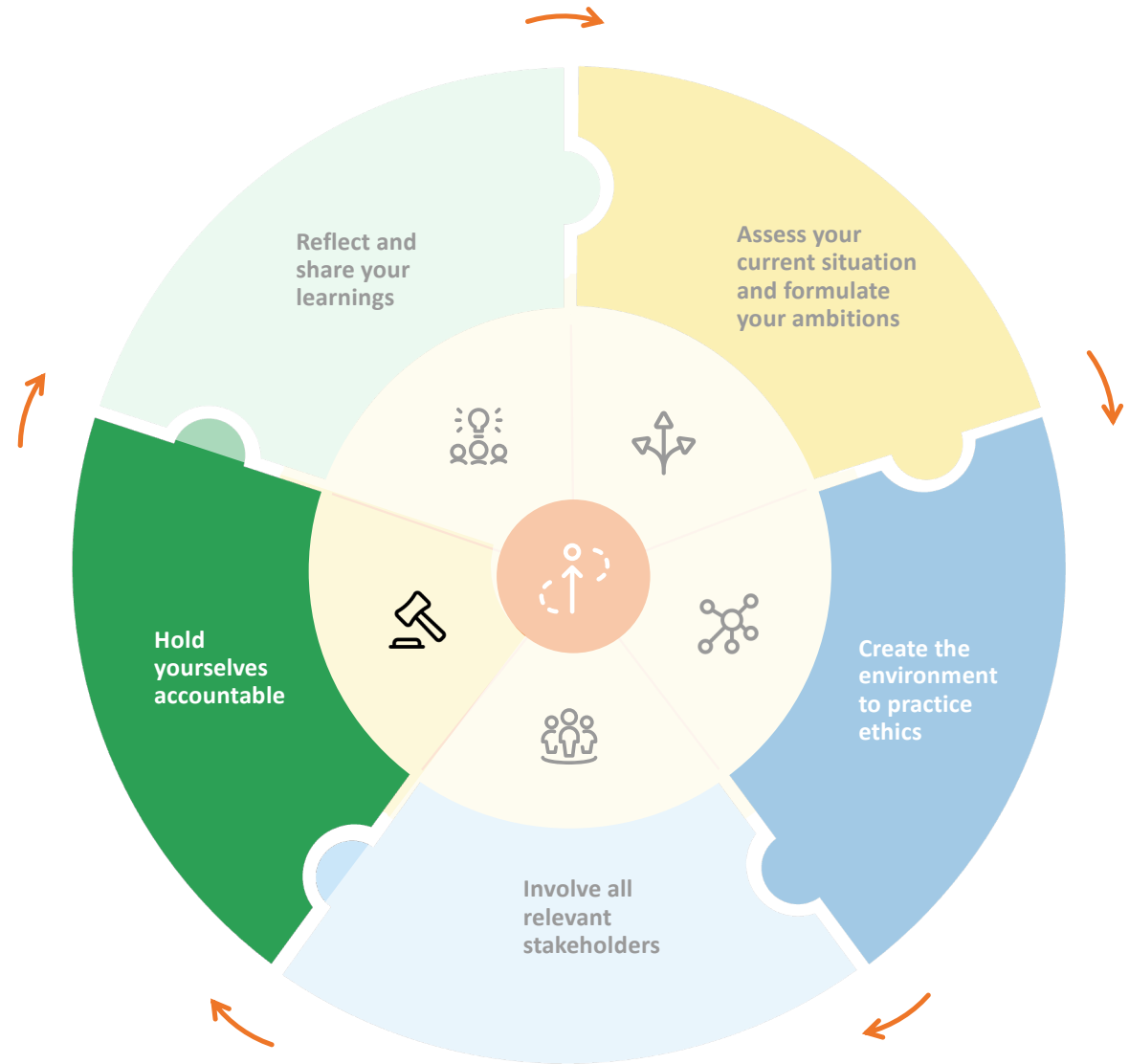
Holding ourselves accountable through transparency

Within our project we commit to (amongst other):

- Publicly communication about commitments, decision workflows, and definitions of success of the project

- All code will be open-source

- Datasheets and model-cards for all datasets and models (according to industry best practices)

- The ambition to publish all our datasets (when possible)

- Training only on data we have licenses for

- Dedication to privacy

- Focus on diversity of the dataset

- Public reporting on our decisions and reasoning

Read (in Dutch still): https://gpt-nl.nl/commitments/

Public commitments help us hold ourselves accountable

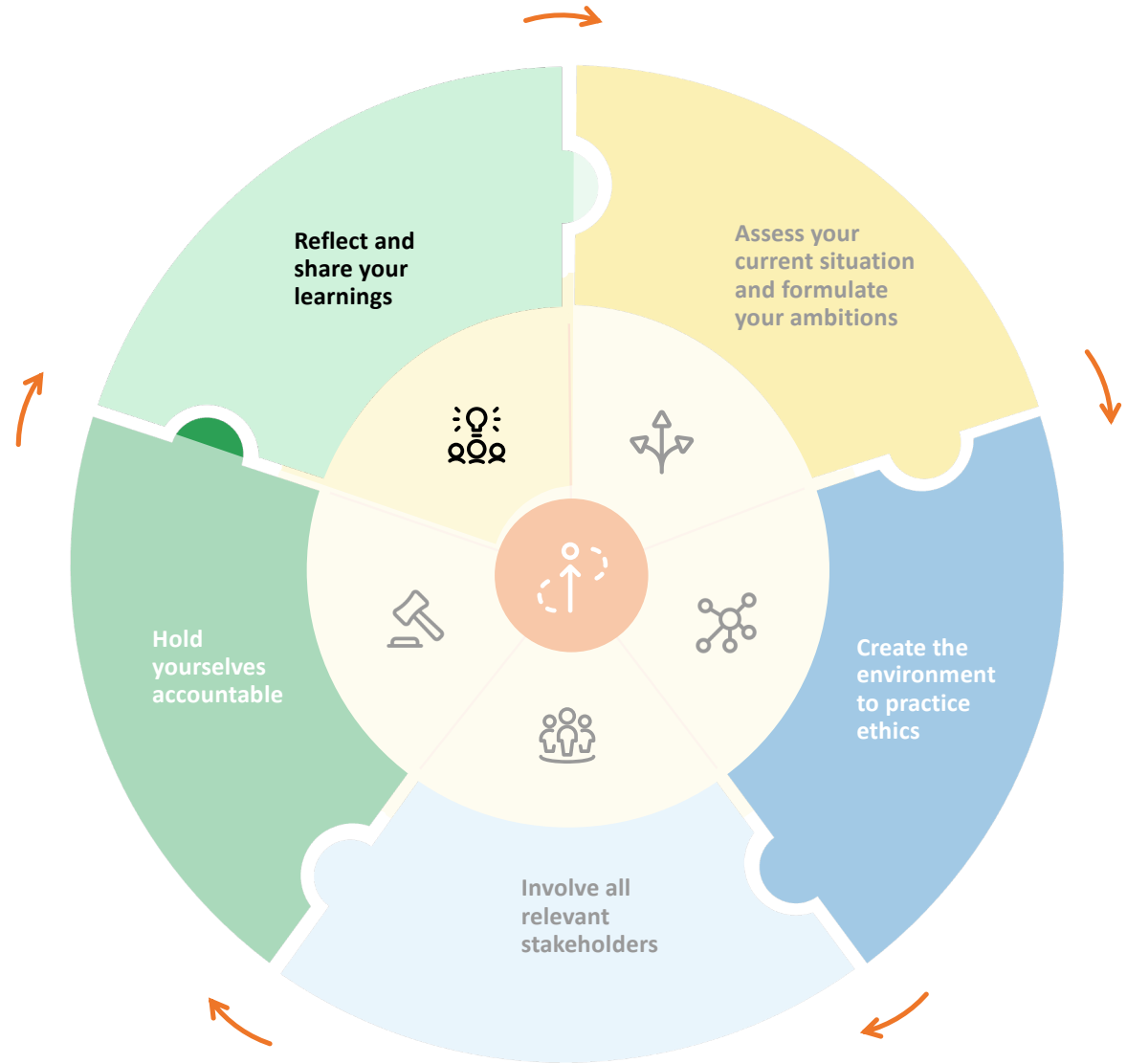Public values can cause tensions

SURF

GPT-NL

How do we get started?

- Create the needed structure to support taking responsibility (e.g. ethics boards)

- Prioritize transparency around technology (e.g. commitments)

- Take responsibility for the applications you promote
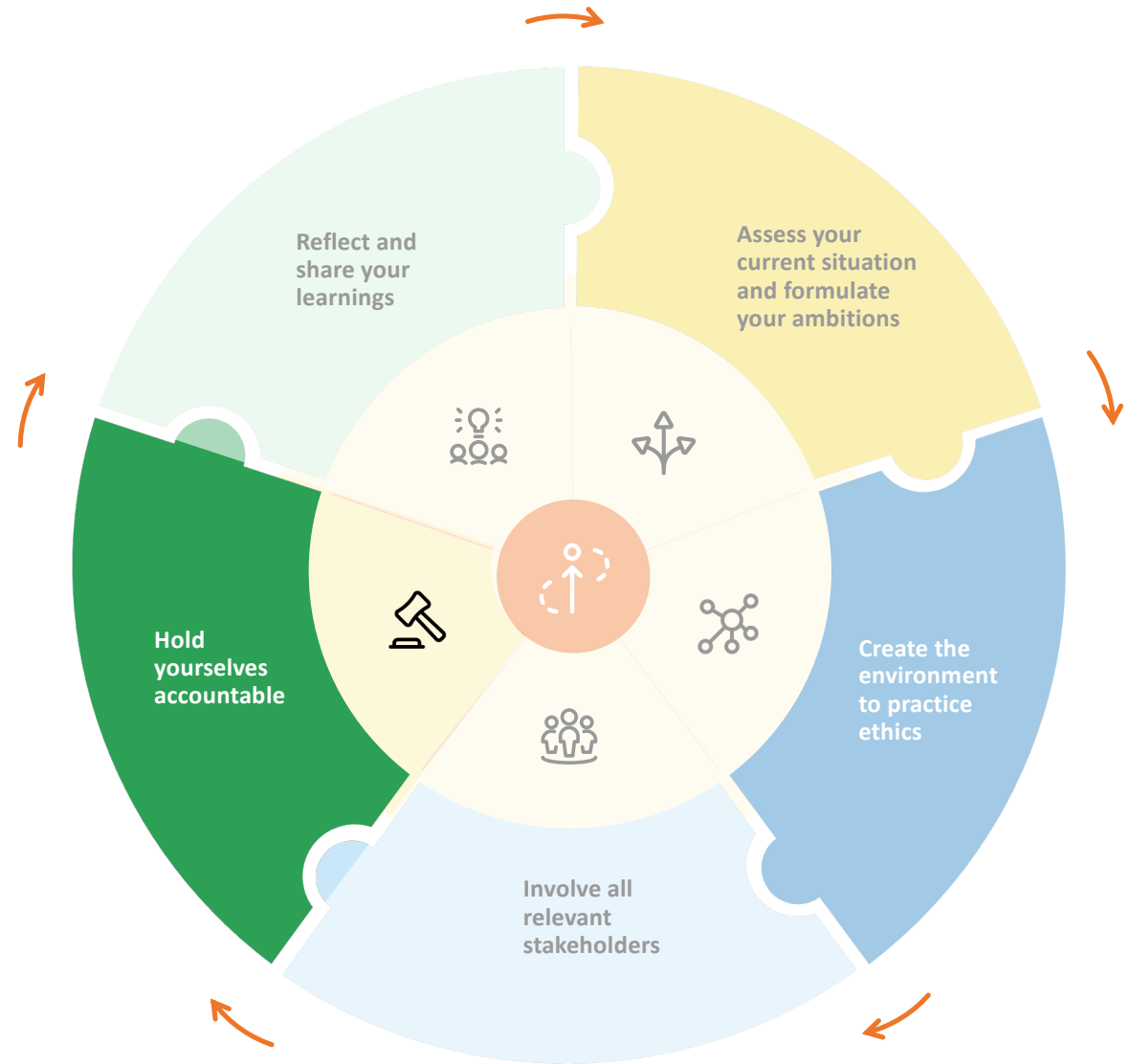
How do we get started?

- Form a learning community, within your organization or with peers

- Share lessons learned or best practices (e.g. collect 'brilliant failures')

Reflect and share your learnings

Assess your current situation and formulate your ambitions

Create the environment to practice ethics

Involve all relevant stakeholders

Hold yourselves accountable

SURF

As long as you're learning, you're not failing.

What does *responsible tech* in the NREN look like?

Let's stop accepting bad technology.
And do better!

Join the Responsible Tech ambition!

Looking forward to hear from you:
duuk.baten@surf.nl

SURF

Links & literature

- Kennisnet. (2020). Weighing Values: an ethical perspective on digitalisation in education. Kennisnet (Authors: Pijpers, R., Bomas, E., Dondorp, L., & Ligthart, J.) ([link](#))

- WRR (2021) 'Opgave AI' - ([link](#))

- Npuls (2024) The state of AI and the modern education institution – ([link](#))

- SURF (2023) Responsible Tech: On Public Values and Emerging Technologies - ([link](#))

- European Commission, Directorate-General for Education, Youth, Sport and Culture, *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2766/153756

- Forthcoming: Molenaar, I., Baten, D., Bard, I., & Stevens, M. (2024). AI and Education: Different Perceptions and Ethical Directions. In N.A. Smuha (Eds.), Cambridge Handbook on the Law, Ethics and Policy of Artificial Intelligence (2024)

SURF